**Social Data Analysis and Visualisation: Preliminary Project Investigations**

* An explanation of the central idea behind your final project (what is the idea? which datasets do you need to explore the idea?, why is it interesting?)

The 2019 - 2020 Coronavirus pandemic will be one of the most influential health crisis of the 21st century. COVID-19, the strain responsible for the pandemic, is a new strain of coronavirus, not previously identified in humans, causing respiratory stress and disease. The health impact continues of the virus is of great concern world-wide. However, the economic impact of the COVID-19 pandemic will continue to be felt for months after the virus has gone. Over the past several weeks, the global share market has dropped, businesses have closed, and governments have been forced to pledge funds to the race to find a vaccine and to help the struggling economy. Governments worldwide have introduced totalitarian social distancing and lockdown laws.

In this project, the link between coronavirus, government and economic data will be drawn upon to compare the response of governments to the pandemic across the Europe. Through the use of coronavirus data including cases by day and country, government data including dates and measures introduced to slow the virus as well as funding pledged in the race to find a vaccine and finally economic data including country debt, GDP, etc, this investigation will analyse how various country indicators have influenced governments responses to the pandemic. Information on the number of hospital beds and healthcare workers will also be sourced to see if there is any correlation between the capacity of the healthcare system and the government measures that have been put into place. This information will be interesting to find links between adopted measures and consequences of the spread of the pandemic. It will also be interesting to see how a countries wealth has impacted the number of coronavirus cases and the government measures that have been enforced.

To explore this idea, economic indicators for example corruption index, GDP and debt will need to be sourced. This information has been retrieved from the world bank data archive. A plethora of data is available from this source and as such, indictors further to the ones listed above have also been found and identified as interesting to the analysis to be conducted. A dataset online containing the government measures introduced worldwide by date was sourced. Finally, coronavirus data, including number of cases, deaths and recovered patients by day was sourced from <https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>. To utilise all of this data, each dataset was individually cleaned and the country codes and names normalised to allow for easy comparison between the datasets. Each dataset contains many attributes and observations. These will need to be filtered and narrowed down to attributes that are relevant and important for this investigation.

* A mock up of the visualization that you wish to build. (Anything is fine here. Pen and paper, MS Paint, Inkscape, D3, anything.).

See below ideas for visualisations, I think if we go ahead with the drawing style video, pen and paper of the mock-up visualisations will be fine. This is also probably the easiest way to do the mock-ups without having to fiddle around with an online tool.

* Make sure you answer the questions
  + What genre is it? (for *Genres*, see section 4.3 of the Segel and Heer paper)
  + Why is that genre right for telling the story you want to communicate with the data

Before diving into the analysis of the economic and government data, it is important to first understand how the COVID-19 pandemic has evolved. This will be shown through a heat map movie over time since the first coronavirus case was confirmed in Europe. This visualisation falls into the film/video/animation genre. As the first visualisation, it is designed to introduce the audience to the story and provide context to the remaining visualisation. The use of the film genre is to provide a light introduction to the story. The visualisation is author driven and as such requires minimal audience participation which, is suitable as an introductory visualisation.

* ‘air\_transport\_carried’ and ‘air\_transport\_worldwide\_departures’ over time versus the number of coronavirus cases
* ‘air\_transport\_carried’ and ‘air\_transport\_worldwide\_departures’ over time since the introduction of lockdown measures (should see a very clear downward trend)
* GDP, debt, tax revenue against the amount of funding pledged by the government
* Visualisation about healthcare expenditure
* Visualisation about the capacity of the healthcare system and the ‘strictness’ of the lockdown (i.e. have some countries been more strict because they don’t have the health system capacity?)
* Smoking, alcohol, diabetes and number of COVID-19 cases
* Handwashing and healthcare expenditure and number of COVID-19 cases
* Government effectiveness and political stability versus number of COVID-19 cases
* Political rights, freedom versus number of COVID-19 cases
* **Number of cases before an individual government put the country in lockdown. (Similar to interactive histogram but with bubbles. Bubble plot, y = no of cases, x = date, colour = country)**
* An outline on the elements you'll need to get to your goal.

To achieve this goal the following elements will be required:

* Data collection
* Data cleaning
* Data reduction
* Aggregation into one large data set
* Initial data explanations
* Initial data analysis
* Visualisation planning
* Creation of video
* Manipulation of data to produce desired visualisations
* Initial production of visualisations
* Editing of visualisations
* Refinement of visualisations
* Presentation of visualisations (titles, labels, etc.)
* Creation of website
* Explanation of motivations, basic statistics, data analysis, genre, visualisations and discussion
* Final editing
* The implementation plan.

These are the required elements for part A:

* Data collection (Luca)
* Data cleaning (Luca)
* Data reduction (Group)
* Aggregation into one large data set
* Initial data explanations (Bethany)
* Initial data analysis (Group)
* Visualisation planning (Bethany)
* Creation of video (Group)

These are the required elements for part B:

* Manipulation of data to produce desired visualisations
* Initial production of visualisations
* Editing of visualisations
* Refinement of visualisations
* Presentation of visualisations (titles, labels, etc.)
* Creation of website
* Explanation of motivations, basic statistics, data analysis, genre, visualisations and discussion
* Final editing
* A walk-through of your preliminary data-analysis, addressing
  + What is the total size of your data? (MB, number of rows, number of variables, etc)
  + What are other properties? (What is the date range? Is is it geo-data?, then a quick plot of locations, etc.)
  + Show the fundamental distributions of the data (similar to the work we did on SF crime data for lecture 3)
* Some general bar plots showing the distributions of the data is all we need, we will probably need to cut down on the data that we have though before we do this.